

# The MIP Ensemble Simulation: Local Ensemble Statistics in the Cosmic Web

M.A. Aragon-Calvo<sup>1</sup><sup>★</sup>

<sup>1</sup>*The Johns Hopkins University, 3400 Charles St., Baltimore, MD, USA.*

Submitted to MNRAS

## ABSTRACT

Here we present a novel N-body simulation technique that allows us to compute ensemble statistics on a local basis, directly relating halo properties to their environment. This is achieved by the use of an ensemble simulation in which the otherwise independent realizations share the same fluctuations above a given cut-off scale. This produces a constrained ensemble where the LSS is common to all realizations while having an independent halo population. By generating a large number of semi-independent realizations we can effectively increase the local halo density by an arbitrary factor thus breaking the fundamental limit of the finite halo density (for a given halo mass range) determined by the halo mass function.

This technique allows us to compute *local ensemble statistics* of the matter/halo distribution at a particular position in space, removing the intrinsic stochasticity in the halo formation process and directly relating halo properties to their environment. This is a major improvement over global descriptors of the matter/halo distribution which can not resolve local variations.

We introduce the *Multum In Parvo* (MIP) constrained ensemble simulation consisting of 220 realizations of a  $32 \text{ h}^{-1} \text{ Mpc}$  box with  $256^3$  particles each. This is equivalent in terms of effective volume and number of particles to a box of  $\sim 193 \text{ h}^{-1} \text{ Mpc}$  of side with  $\sim 1540^3$  particles containing  $\sim 5 \times 10^6$  haloes with a minimum mass of  $3.25 \times 10^9 \text{ h}^{-1} \text{ M}_\odot$ . The MIP simulation stands apart from all previous N-body simulations in its unprecedented high equivalent particle density and local halo density.

We illustrate the potential of the technique presented here by computing the *local mass function* at several characteristic environments and along a path from the center of a void to its border. We can study for the first time the effect of local environment in the height, shape and characteristic mass of the halo mass function.

Future observations will provide detailed models of the galaxy density distribution allowing the use of constrained realizations in combination with ensemble simulations to derive accurate “mean ensemble” properties of the local Universe.

**Key words:** Cosmology: large-scale structure of Universe; methods: data analysis, N-body simulations

## 1 INTRODUCTION

N-body simulations are one of our most valuable tools in cosmology. They allow us to follow the evolution of cosmic structures from the linear regime to the present time over a wide range of scales and masses. Current state-of-the-art N-body simulations contain billions of particles and can resolve from large structures like superclusters and clusters of galax-

ies down to subhaloes and dwarf galaxies (Springel et al. 2005; Diemand et al. 2007; Springel et al. 2008).

One area where N-body simulations have been particularly useful is in the study of the formation and evolution of haloes and their relation with their large-scale environment. Haloes are affected by their environment through the local density, tidal field, matter accretion, mergers, etc. (White (1984); Byrd & Valtonen (1990); Lacey & Cole (1993); Zabludoff & Mulchaey (1998); van den Bosch (2002); Gottlöber et al. (2001) among many

\* E-mail:miguel@pha.jhu.edu

## 2 Aragon-Calvo M.A.

others). In order to properly understand the role of local environment in the process of galaxy formation/evolution it is crucial to be able to perform *local* correlations between haloes and their local large-scale structure (LSS). An important application of this is the use of simulations to interpret real observations of specific (and often complex) environments such as our local neighborhood. This is an non-trivial task as the many “problems” existing between observations and simulations show. Examples of such discrepancies include the “cold Hubble flow” problem (Sandage et al. 1972; Sandage 1986), the “local velocity anomaly” (Tully et al. 2008), the excess of luminous galaxies above the plane of the local sheet (Peebles & Nusser 2010) and, to some degree, the “missing satellite” problem (Klypin et al. 1999).

Until recently our current set of theoretical and numerical analysis tools was limited on global descriptors such as the two-point correlation function and its generalization to  $n$ -points (Peebles 1980; Szapudi 1998), marked correlations (Sheth 2005), topology via the Minkowski functionals (Mecke et al. 1994; Schmalzing et al. 1999), etc. These tools provide a good description of the global clustering and scaling relations of the matter in the Universe. However, since such measurements contain no phase information they are insensitive to *local* variations that may arise in the diversity of environments in the Cosmic Web. Some important advances have been done in this respect with the introduction of LSS characterization algorithms based properties of the LSS such as the density field (Aragón-Calvo et al. 2007) and related Hessian-based descriptors (Zhang et al. 2009; Cautun et al. 2012), tidal field (Hahn et al. 2007), shear tensor (Libeskind et al. 2012), topology (Sousbie et al. 2008; Aragón-Calvo et al. 2010) among many others. Using such tools haloes can be assigned to particular environments and then study the dependence of their properties. However, after the assignment of haloes of specific environments, invariably the halo sample is integrated into global environment-based samples, thus losing the locality gained by the LSS characterization. The main reason for this is the need for a large halo sample in order to perform meaningful statistical analysis. This would be in principle sufficient if all environments could be completely characterized using the algorithms mentioned above. However, even the most complete characterization tools available can not fully describe the complexity of environments found in the Cosmic Web.

### 1.1 Finite halo sampling

The problem of correlating halo properties with their particular environment can be then described as a “sampling problem”. Local studies of haloes and their environment are ultimately limited by the local density of haloes which depends on the halo mass given by the halo mass function (Press & Schechter 1974). The halo mass function sets limits to the number of haloes of a given mass that we expect to find in a given environment. We refer to this as the *specific halo number density*. As the halo mass increases their specific number density decreases until there are not enough haloes to perform any statistical analysis or even until there are no haloes at all as in the case of the regions at the center of voids (Gottlöber et al. 2003). In the Universe and standard computer simulations we only have one realization to make measurements. At any point in space and at a given

halo mass range there is a predefined halo number density given by the halo mass function. The specific halo density does not depend on the mass resolution of our simulation. A higher density of particles will only increase the number of lower-mass haloes. This sets fundamental limits to the kind of analysis we can perform on a local basis. One could for instance measure statistics of haloes inside a sampling window small enough to be sensitive to local variations. While this may work for very limited cases such as low-mass haloes and high density regions, in general one simply does not have enough haloes to obtain statistically significant results. This effect is particularly severe in low density regions such as voids and walls where the density of halos massive enough to host luminous galaxies is of the order of one halo per several cubic megaparsec.

In this paper we present a new technique that solves the problem of a finite specific halo density. The technique is based on a constrained ensemble simulation where all semi-independent realizations share the same large-scale fluctuations. Ensemble simulations are becoming an important tool to study the statistical properties of the matter/galaxy distribution where a large sampling volume is required (Meiksin & White 1999; Takahashi et al. 2009; Smith 2009; Schneider et al. 2011; Forero-Romero et al. 2011; Orban 2012). For instance, Takahashi et al. (2009) ran 5000 independent realizations of a  $1 \text{ h}^{-1} \text{ Gpc}$  box with  $256^3$  particles each to study the Baryon Acoustic Oscillations (see also Schneider et al. (2011)). At smaller physical scales, Forero-Romero et al. (2011) used ensemble simulations from constrained random realizations of the local supercluster in order to identify structures of interest. These authors ran 200 constrained realizations of the local supercluster and identified “local group” systems in only 3 realizations. From these 3 realizations higher resolution simulations were generated and analyzed. As far as we know the remaining 197 realizations in the ensemble were not used.

## 2 LOCAL ENSEMBLE SIMULATIONS

The overall structure and dynamics of the Cosmic Web is mainly determined by large-scale fluctuations with smaller fluctuations playing only a minor role in the overall picture (Little et al. 1991; Suhonenko et al. 2011; Einasto et al. 2011a). Since the primordial density field is a linear combination of fluctuations it can be separated into two regimes: one responsible for shaping the large-scale features of the Cosmic Web and another with smaller galaxy-size fluctuations as:

$$\delta = \delta(< k_{\text{cut}}) + \delta(\geq k_{\text{cut}}) \quad (1)$$

where  $\delta$  is the total primordial density field,  $\delta(< k_{\text{cut}})$  is the large-scale fluctuations field:

$$\delta(< k_{\text{cut}}) = \delta(k) W(k). \quad (2)$$

and  $\delta(\geq k_{\text{cut}})$  is the high-pass filtered density field given by:

$$\delta(\geq k_{\text{cut}}) = \delta(k) (1 - W(k)). \quad (3)$$

Where  $W(k)$  is a filter with a cutoff frequency  $k_{\text{cut}}$  given by:

$$W(k) = \begin{cases} 0, & k > k_{\text{cut}} \\ 1, & k \leq k_{\text{cut}} \end{cases} \quad (4)$$

Following equation 1 we can generate multiple semi-independent realizations from the same primordial density field by fixing the large-scale fluctuations  $\delta(< k_{\text{cut}})$  between realization and allowing the small-scale fluctuations  $\delta(\geq k_{\text{cut}})$  to vary. The choice of  $k_{\text{cut}}$  determines the scale and equivalent mass below which realizations are independent between each other:

$$M(k_{\text{cut}}) = \frac{4}{3}\pi \left( \frac{L_{\text{box}}}{2\pi k} \right)^3 \bar{\rho}. \quad (5)$$

Density fluctuations below this scale are statistically independent at the initial conditions and latter become correlated by the power transfer from large scales. These semi-independent realizations define multiple paths for initially statistically independent small-scale objects that nevertheless converge to the same large-scale patterns by their shared large-scale fluctuations. Haloes with masses larger than  $M(k_{\text{cut}})$  are not independent between relations offering a unique opportunity to study their “persistent” properties across the ensemble. These properties reflect the large-scale environment of the halo as the small-scale contributions are averaged-out by the ensemble.

## 2.1 Nested ensembles

The idea of a constrained ensemble can be extended to iteratively generate nested constrained ensembles where for each realization  $i$ :

$$\delta^i = \delta(< k_{\text{cut}}) + \delta(\geq k_{\text{cut}})^i \quad (6)$$

we generate a new ensemble defined by a new cut-off scale  $k_{\text{cut}2}$  where  $k_{\text{cut}2} > k_{\text{cut}}$ . Realizations in the nested ensemble will share structures at scales larger than  $k_{\text{cut}2}$  and will be independent below those scales:

$$\delta_2^i = \delta(< k_{\text{cut}}) + \delta(\geq k_{\text{cut}}, < k_{\text{cut}2})^i + \delta(\geq k_{\text{cut}2}). \quad (7)$$

Here the subscript denotes the level of recursion in the ensemble.

Designing a nested constrained ensemble one could define a cutoff scale  $k_{\text{cut}}$  corresponding to cluster-group size structures and having a galaxy-size independent halo population between realizations. The next recursion level would be then determined by a higher  $k_{\text{cut}2}$  corresponding to Milky-Way size galaxies thus generating a sub-halo independent population between realizations for every galaxy-size object in the first level of the ensemble.

The number of realizations grows with the number of nested ensembles as  $N^m$  where  $N$  is the number of realizations per ensemble and  $m$  is the level of recursion. This exponential behaviour makes it unfeasible to generate more than a just few recursion levels. In the present work we present a one-level ensemble. A two-level nested ensemble is in preparation at the time of writing.

## 2.2 Practical advantages

The self-contained nature of each realization offers several practical advantages over standard N-body simulations:

- Trivial to parallelize (run and post-processing). Realizations are self-contained and can be run and analysed on an independent pipeline.

- Computing time increases linearly with the number of realizations.
- Adding more realizations is trivial as it only involves generating new small-scale fluctuations.
- No need for custom read-write routines as in the case of standard massive simulations where parallel IO routines are necessary given the shear size of the datasets.
- Efficient storage of many relatively small snapshots.

It may seem that the above advantages are the same as if we simply ran a large number of independent realizations. However, in our case each realization corresponds to the same LSS configuration so all realization can be effectively considered to be the same simulation with different halo formation paths.

## 3 LOCAL ENSEMBLE STATISTICS

The concept of ensemble implies the existence of statistically independent events. The ensemble average of a variable of the system is given by

$$\langle w \rangle = \sum_{i=1}^N w_i p_i \quad (8)$$

where  $w_i$  is the measured variable at event  $i$  and  $p_i$  is the probability of observing the event  $i$  with  $p_i = 1/N$  when all events have the same probability of being “observed”. Angled brackets denote the ensemble average. In general it is desired to know the moments of a variable with respect to its ensemble average. The  $n^{\text{th}}$  central moments of a distribution are then:

$$\langle |w - \langle w \rangle|^n \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (w_i - \langle w \rangle)^n. \quad (9)$$

where  $\langle w \rangle$  is the ensemble average defined in equation 8 and  $N$  is the number of realizations or observations in the ensemble. It is common to invoke ergodicity and interchange volume averages by ensemble averages. In our case however, since realizations in the ensemble are correlated above  $k_{\text{cut}}$  this is not necessary and we can define the *local ensemble moments* at position  $\mathbf{x}$  as:

$$\langle |w(\mathbf{x}) - \langle w(\mathbf{x}) \rangle|^n \rangle = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N (w(\mathbf{x})_i - \langle w(\mathbf{x}) \rangle)^n. \quad (10)$$

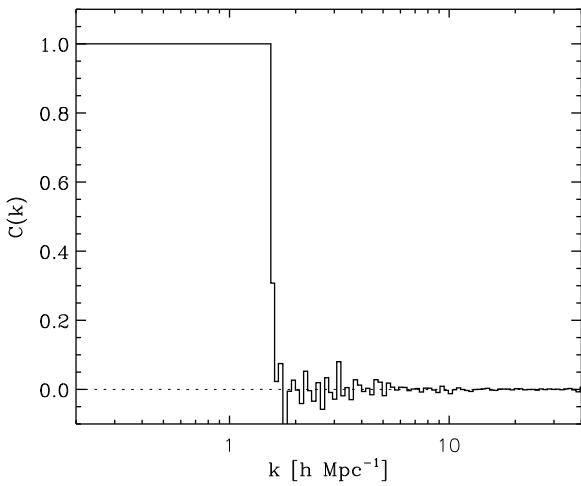
In practice we compute the local  $n^{\text{th}}$  central ensemble moment by considering all  $M$  sampling points (halos or particles) inside a sampling window of radius  $R$  located at position  $\mathbf{x}_i$  across all  $N$  realizations:

$$\langle |w(\mathbf{x}) - \langle w(\mathbf{x}) \rangle_R|^n \rangle_R = \frac{1}{N} \sum_{j=1}^N \frac{1}{M} \sum_{i=1}^M (w(\mathbf{x}_{i,j}) - \langle w(\mathbf{x}) \rangle_R)^n. \quad (11)$$

where  $w(\mathbf{x}_{i,j})$  is the variable measured at sampling point  $\mathbf{x}_i$  inside the sphere corresponding to realization  $j$  and  $\langle w(\mathbf{x}) \rangle_R$  is the local ensemble mean.

### 3.1 Realization stacking and signal-to-noise

The cross-correlated nature of the realizations in the constrained ensemble allows us to “stack” them into one single



**Figure 1.** Cross-correlation coefficient between the initial conditions of two realizations in the ensemble.

field. By doing so we ‘average-out’ the effect of stochastic processes and focus on the effect of the environment defined by scales larger than the cut-off scale  $k_{\text{cut}}$ . This is a similar process than the “image stacking” technique used in image processing for noise reduction. The signal-to-noise ratio of the stacked ensemble is given by:

$$s/n = \frac{N}{\sqrt{N}}. \quad (12)$$

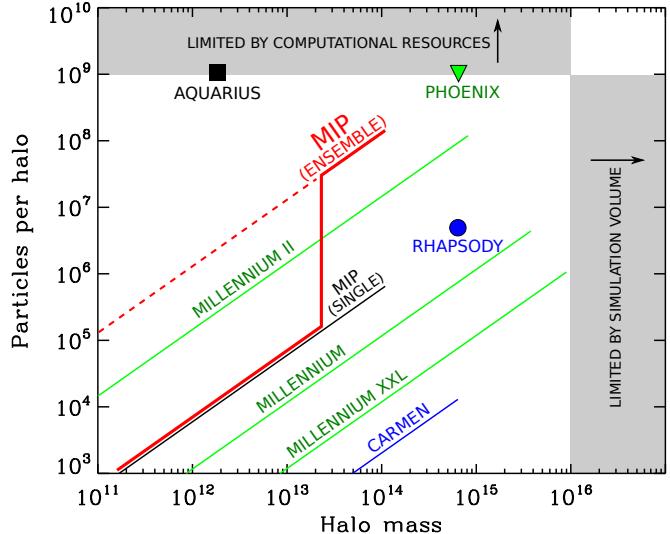
Where  $N$  is the number of realizations in the ensemble. Although in practice we can not create an infinite number of realizations one can generate enough realizations in order to increase the signal-to-noise ratio to any desired level.

#### 4 THE MIP SIMULATION: SETTING UP

Here we describe the *Multum In Parvo*<sup>1</sup> (MIP) ensemble simulation. The MIP simulation is an undergoing project consisting (in its first stage) of 512 realizations of a  $32 h^{-1}\text{Mpc}$  box sampled on a  $256^3$  regular grid giving a mass per particle of  $1.62 \times 10^8 M_\odot h^{-1}$ . The cosmology corresponds to the standard  $\Lambda\text{CDM}$  with parameters  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ ,  $h=0.73$ ,  $\sigma_8 = 0.84$  and spectral index  $n = 0.93$ .

The results presented here correspond to 220 out of a total of 512 realizations that have been completely run and analyzed to identify FoF haloes in all snapshots and Sub-Find subhaloes at the final snapshot as described in Section 4.2. The full MIP ensemble simulation is equivalent in terms of volume and number of particles to a  $256 h^{-1}\text{Mpc}$  box standard simulation with  $2048^3$  particles. The current status of 220 realization is equivalent to a  $32 \times (220)^{1/3} \sim 193 h^{-1}\text{Mpc}$  box with  $256 \times (220)^{1/3} \sim 1540^3$  particles. With 220 realizations the signal-to-noise ratio is  $s/n = \sqrt{220} \simeq 15$ . A comparison between the properties of a single realization and the ensemble simulation is presented in Table 1.

<sup>1</sup> Latin for “much in little”. The expression is also used in computer graphics to denote image pyramids (mipmaps) for texture rendering.



**Figure 2.** Comparison of the number of particles per halo mass between the MIP and several simulations in the literature. Lines indicate simulations with a “continuous” halo population and isolated symbols indicate a resimulation of individual halos. The single-realization MIP is shown as a black line. The MIP ensemble simulation is shown in red. The lower regime corresponds haloes with masses below  $M(k_{\text{cut}})$ . In this mass regime haloes are independent between realizations and can not be stacked into an “ensemble halo”. The upper regime of the MIP simulation corresponds to haloes with masses larger than  $M(k_{\text{cut}})$ . In this mass regime haloes can be stacked as they are cross-correlated across the ensemble. Note that this diagram does not refer to the specific halo density (see text for details) which is 220 times larger for the MIP compared to standard simulations. The MPI simulation occupies a unique region in the diagram in terms of its two-regime behaviour and unprecedented particle count for a continuous halo population. It has the highest number of particles per halo above  $M_{\text{cut}}$ .

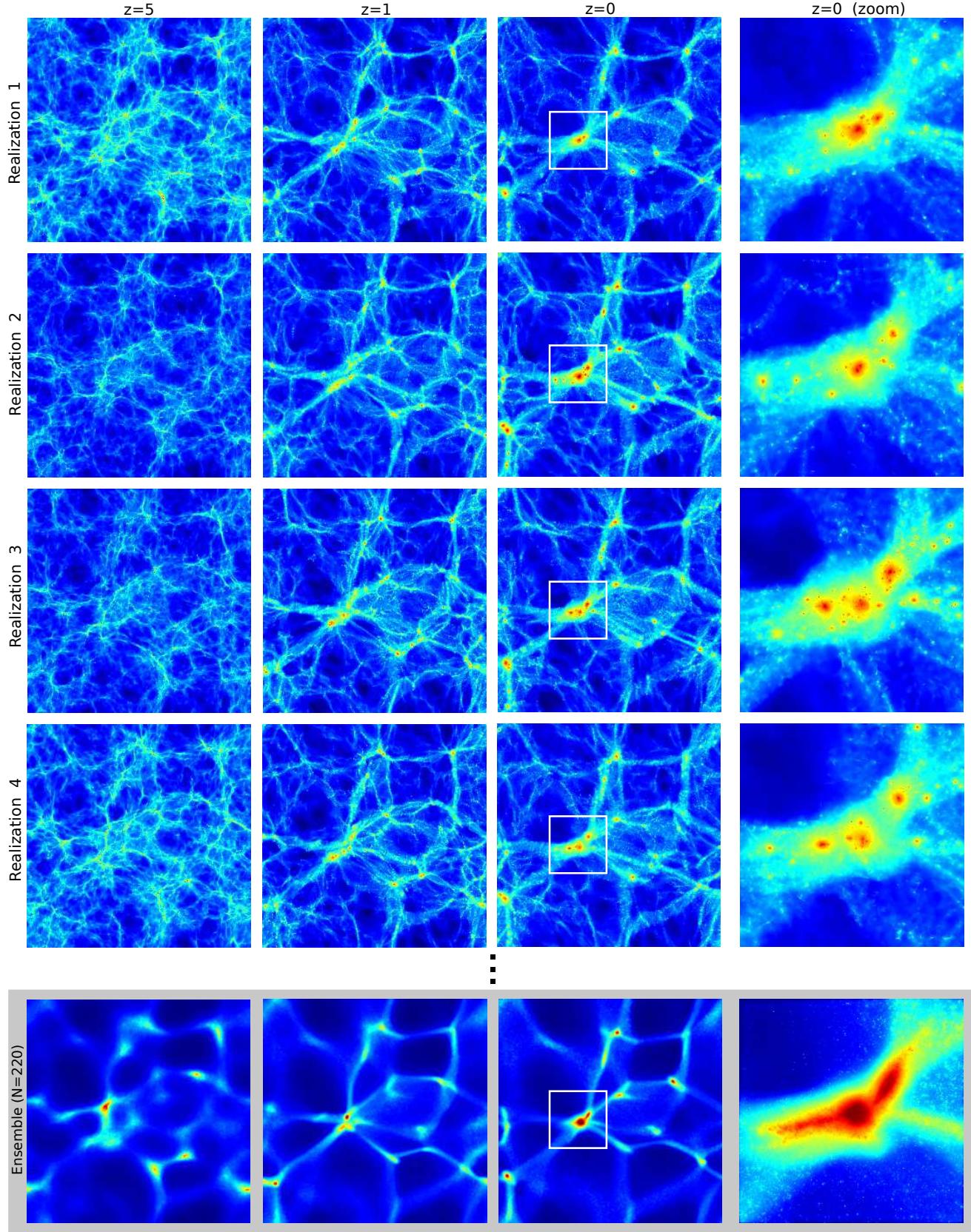
#### 4.1 Generating initial conditions

The first step in making the constrained ensemble is the creation of a “template” initial conditions field from which all realizations will be generated. The template field was generated using the publicly available `GRAFIC` code. From the template each of the semi-independent realizations was generated as described in section 2 (see also Figure A1). In practice we do not generate a new random field for each realization but instead shift  $\delta(> k_{\text{cut}})$  in real space along each dimension by a distance corresponding to  $k_{\text{cut}}$ . For our study we set  $k_{\text{cut}} = 4 h^{-1}\text{Mpc}$  corresponding to a volume containing a mass of  $\sim 2 \times 10^{13} h^{-1} M_\odot$ . For the  $32 h^{-1}\text{Mpc}$  box this gives  $32/4 = 8$  independent shifts per dimension for a total of  $8 \times 8 \times 8 = 512$  realizations. From the set of initial condition files we then generated Gadget files starting at  $z \simeq 80$  following the standard Zel'dovich prescription (Zeldovich 1970).

Figure 1 shows the cross correlation coefficient between two initial density field realizations computed as:

$$C_{1,2}(k) = \frac{P_{1,2}(k)}{\sqrt{P_1(k) P_2(k)}}. \quad (13)$$

There is a perfect cross-correlation between the two realizations at scales  $< k_{\text{cut}}$  as expected since by construction both



**Figure 3. Ensemble simulation.** Four realizations are shown at redshifts  $z = 5, 1$  and  $0$  (first three panels from left to right). The stacking of the 220 realizations in the ensemble is shown in the bottom panels. The colors indicate the local density on a thin slice across the simulation box. A zoom region of  $8 \times 8 h^{-1}$  Mpc centered on a cluster is shown in the right panels.

initial density fields share the same large-scale fluctuations. At scales ( $k > k_{\text{cut}}$ ) there is practically no cross-correlation although there are small noisy “ripples” possibly arising as a result of having the same small scales only shifted instead of being completely independent realizations. Figure 1 shows that while small-scale fluctuations from adjacent realizations produced by shifting are not strictly independent from each other they are sufficiently separated to minimize significant cross-correlation.

#### 4.1.1 Ensemble run

Each realization was ran on 16 processors on the Homewood High Performance Cluster (HHPG) at JHU. We stored 50 snapshots separated in logarithmic intervals in scale factor starting at  $z = 5$ . The realizations were ran and processed in batches of 20 runs in order to keep a low load on the HHPG cluster. All the snapshots from the 220 realizations plus halo/subhalo catalogues occupy approximately 5.5 TB. Each “ensemble snapshot” occupies  $\sim 100$  GB in contrast to the much smaller “realization snapshot” with only  $\sim 460$  MB.

#### 4.1.2 MIP compared to standard simulations

Figure 2 shows a comparison between the MIP and several other simulations in the literature: Carmen (from “Las Damas”, McBride et al. 2011, in prep), Millennium (Springel et al. 2005), Millennium II (?) and Millennium XXL (Angulo et al. 2012). We also show three high-resolution zoom resimulations: Aquarius project (Springel et al. 2008), Phoenix (Gao et al. 2012) and Rhapsody (Wu et al. 2012). This is by no means a complete list. It is intended to show the current state-of-the-art in N-body cosmological simulations. The single-realization MIP behaves like a standard simulation. It is, for current standards, a small-box medium-resolution simulation. It sits between the Millennium and Millennium II simulations in terms of particle resolution but it contains a much smaller volume. On the other hand, the MIP ensemble simulation occupies a unique region in the diagram. It has the highest number of particles per halo in its “local ensemble range” determined by  $M(k_{\text{cut}})$  (almost one order of magnitude higher than millennium II). Haloes in this regime are persistent across the ensemble with small variations between realizations allowing us to study their ensemble statistics. Haloes less massive than  $M_{\text{cut}}$  are independent between realizations and can be stacked to increase the specific halo density. In our case the density of haloes is 220 times larger than in the single-realization case. For comparison, the density of *haloes* in the MIP simulation is almost two times higher than the density of *particles* in the millennium I simulation! Given its unique properties, one can not directly compare the MIP ensemble simulation to other standard simulations. The MIP simulations stands apart from standard simulations in terms of particles per halo, specific halo density, statistical properties, and computational requirements.

## 4.2 Halo/subhalo catalogues

We identified Friends of Friends (FoF) haloes for all 50 snapshots of each realization using a linking parameter of  $b = 0.2$ . Only haloes with a minimum mass of 20 particles were used. For each halo we computed extensive physical properties such as mass, radius, inertia tensor, angular momentum,  $v_{\text{max}}$ , etc. The details of the computation of halo properties will be described in detail in a forthcoming paper. Subhaloes were identified only at the final snapshot at  $z = 0$  with the SUBFIND code (Springel et al. 2001) which identifies subhaloes as gravitationally bounded substructures inside FoF haloes. We also computed physical properties for the subhaloes and stored both halo and subhalo catalogues in a database for efficient retrieval.

Given the large number of realizations, the creation, running and analysis of each of the 512 realizations is controlled with an automated pipeline. The final result is a set of halo/subhalo catalogues and most massive progenitor lists for FoF haloes.

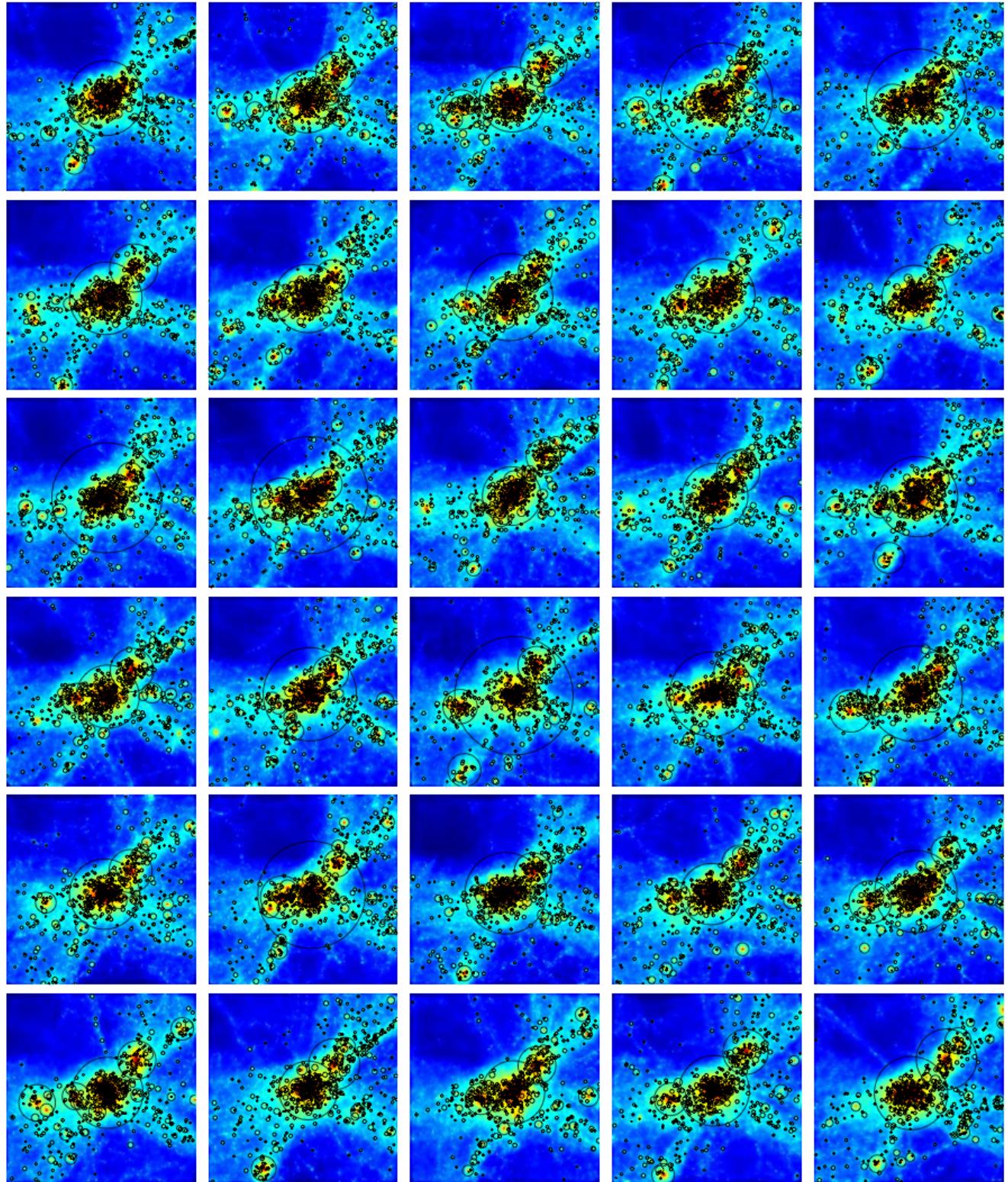
## 5 THE MIP SIMULATION: FIRST RESULTS

We now describe the general properties of the MIP simulation starting with individual realizations, the global mass function (per realization and ensemble) and the basic properties of the most massive haloes in the simulation.

### 5.1 Individual realizations

Figure 3 shows an overall view of the MIP ensemble simulation. Four individual realizations are shown as well as the stacked ensemble. From left to right we show the density field at redshifts  $z = 5, 1$  and  $0$  and a zoom into the central halo in the slice. All realizations share the same large-scale features as expected since they evolved from the same large-scale initial fluctuations. Already at  $z = 5$  we can identify common structures between realizations like the large void at the top-left corner and the region that will collapse into the largest halo at the center of the slice. The small-scale fluctuations on the other hand are different in each realization. This becomes more evident as the simulation evolves and the fluctuations collapse into haloes with unique substructure. At  $z = 1$  it is clear that even when the LSS is roughly the same between realizations it is composed by a unique halo population. This effect can be seen in all environments but it is perhaps more dramatic in walls and voids where the halo density is very low and often the same region is unevenly sampled between realizations.

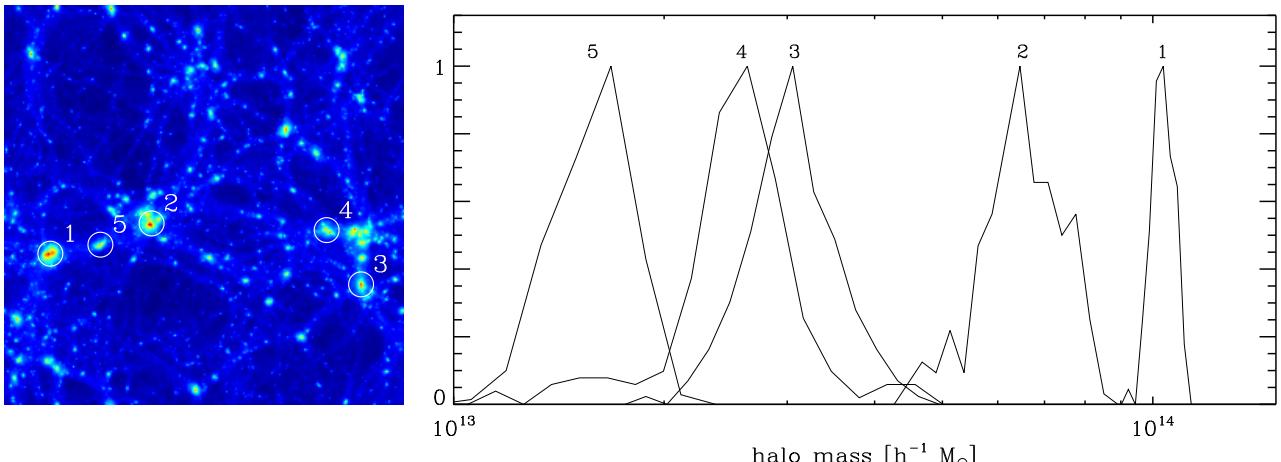
All realizations produce a similar cluster at the center of the slice. However, a detailed view inside the cluster shows significant differences at small scales. For instance, the central cluster in the first two realizations of Figure 3 has two large subhaloes while the last two haloes have three large subhaloes. The differences at smaller scales are even larger making it impossible to identify common low-mass subhaloes between realizations. The halo population corresponding to these scales is effectively independent between realizations. Despite this, low-mass haloes trace the same large-scale structures as a result of the transfer of power from large scales to small scales via gravitational coupling.



**Figure 4.** Thirty realizations of the central cluster in Figure 3. The SubFind subhaloes are shown as open circles scaled with their radius. The background image represents the dark matter density field for the corresponding realization.

	$V_{\text{box}} [\text{h}^{-1} \text{Mpc}]^3$	$V_{\text{eq}} [\text{h}^{-1} \text{Mpc}]^3$	$N_{\text{part}}$	$N_{\text{halos}} (z = 0)$	$s/n$
single	$32^3$	$32^3$	$256^3$	$\sim 2 \times 10^4$	1
ensemble	$32^3$	$193^3$	$1545^3$	$\sim 5 \times 10^6$	$\sim 15$

**Table 1.** MPI ensemble simulation, comparison between single and ensemble ( $N=200$ ) properties.  $V_{\text{box}}$  corresponds to the volume of the simulation box and  $V_{\text{eq}}$  the equivalent volume of the ensemble.  $N_{\text{part}}$  is the total number of dark matter particles.  $N_{\text{halos}}$  is the number of haloes above the minimum particle limit of 20 particles corresponding to  $3.2 \times 10^9 \text{ h}^{-1} \text{ M}_\odot$ .  $s/n$  is the signal-to-noise ratio of the stacked ensemble.



**Figure 5. Five largest haloes in the simulation.** The left panel indicates the location of the haloes in the simulation box with white circles. The right panel shows the mass distribution for each halo across all realizations in the ensemble. The ensemble mean mass of the most massive halo in the simulation is  $1.0 \times 10^{14} \pm 3.8 \times 10^{12} \text{ h}^{-1} \text{ M}_\odot$ .

## 5.2 Stacked ensemble

The ensemble density field is shown in the bottom panels of Figure 3. This field was created by stacking the individual realizations in the ensemble. The same idea is used in image processing in the ‘‘image stacking’’ technique where a sequence of images taken from the same subject are added in order to reduce the noise level (see Section 3.1). In our case the stacking of realizations ‘‘averages out’’ random fluctuations of scales smaller than  $k_{\text{cut}}$ .

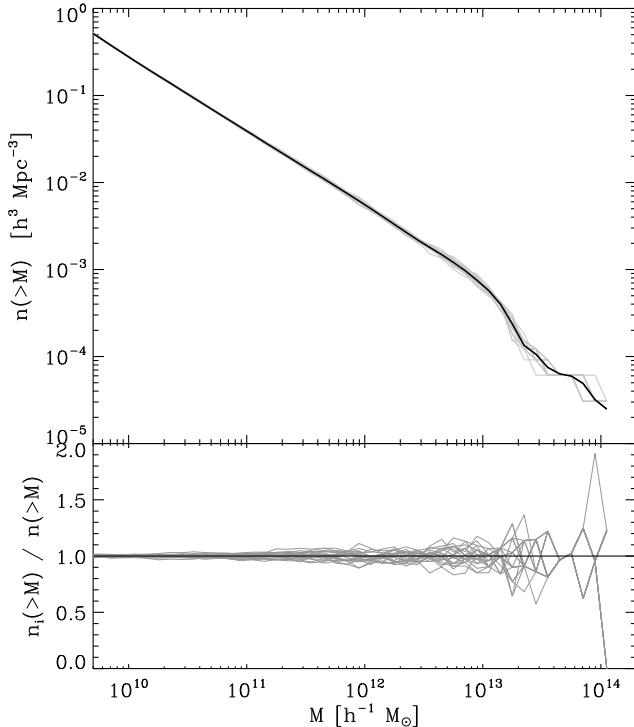
The stacked density field has some interesting properties that make it useful for applications where a smooth field is needed. The stacking of realizations acts like a sort of low-pass filter producing a smooth density field while at the same time retaining anisotropic features, in contrast to what we would expect if the field was smoothed with a Gaussian kernel. Note that in contrast to the single realizations, the ensemble density field looks remarkably similar at all redshifts. The only changes we can see seem to be due to scales larger than  $k_{\text{cut}}$  as smaller scales are averaged by the stacking. The ensemble density field can be used as a robust large-scale density field tracer which is not affected by small scale fluctuations such as haloes which tend to negatively affect structure finder techniques based on the density field. A good illustration of this is the thin filament located at the bottom-right of the central cluster. In all four realizations there are significant differences in the substructures defining the filament. The ensemble density field, on the other hand, shows a well defined tenuous structure with practically no substructure and an almost constant density profile.

The ensemble density field looks tantalizing similar

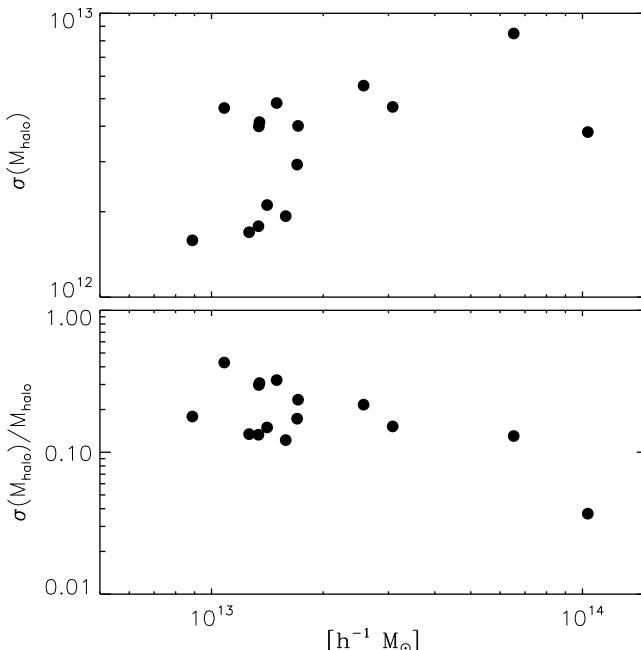
to a Lagrangian-smoothed density field (see (Little et al. 1991; Melott & Shandarin 1993; Avila-Reese et al. 2001; Einasto et al. 2011b; Aragon-Calvo et al. 2010, 2012)) where the primordial density field is smoothed in the linear regime and then evolved. Such Lagrangian-smoothed simulations contain only large-scale modes and therefore no substructure below the smoothing scales. The top-right panel of Figure 8 shows the lagrangian-smoothed version of the cluster in Figure 8. Note the similarity with the cluster in the ensemble density field shown in the same figure and in the bottom-right panel of Figure 3. The Lagrangian-smoothed density field presents several artifacts from the density reconstruction procedure but in general it has the same properties than the ensemble density field. However, the lagrangian-smoothed version does not contain any small-scale structure and therefore can not be directly used to study haloes.

## 5.3 Ensemble halo population

Figure 4 shows 30 versions of the central cluster in Figure 3 identified in different realizations. The ensemble mean mass of the central FoF halo is  $\sim 10^{14} \text{ h}^{-1} \text{ M}_\odot$ , computed as the mean mass of the corresponding halo across the ensemble. This is the second largest halo in the ensemble. We highlight the substructure inside the cluster by the position of the subhaloes as open circles scaled with their radius. While the megaparsec-scale matter distribution is roughly similar in all realizations there is a large variety in the properties of the halo/subhalo populations. This can be seen in the



**Figure 6. Halo mass function.** Top: the black solid line corresponds to the halo mass function of the entire ensemble. The thin gray lines show the halo mass function computed from 20 individual realizations. Bottom: ratio between the mass function of each realization and the ensemble mass function.



**Figure 7.** Top: Halo mass dispersion for the 15 most massive haloes in the simulation. Bottom: relative halo mass dispersion.

most massive subhalo of the cluster which has a variation in its size of almost a factor of two between the most extreme versions of the cluster shown in the figure. Each realization of the cluster shows a unique formation path with different mass and halo accretion histories, internal dynamics etc. However, because in all realizations the same LSS is shared one expect the properties of the cluster to be highly correlated across the ensemble (Araya-Melo & Aragon-Calvo in preparation).

### 5.3.1 Ensemble Halo Mass Dispersion

In order to get a better understanding of the differences between halo populations in the ensemble we computed the ensemble halo mass dispersion for the 15 most massive haloes in the simulation. We do this by first defining a “reference” sample of haloes identified from a single realization. Next we track the reference haloes across all realizations in the ensemble. We track a given halo by placing a search window at the halo’s position in all realizations and selecting the most massive halo inside the search window. The search window radius is of the same size as the original halo. Figure 7 shows the ensemble mass dispersion for the 15 most massive haloes. The relative dispersion is shown in the lower panel. We normalized the dispersion by the ensemble mean halo mass for each halo. The relative dispersion increases for decreasing halo mass. The trend in the relative dispersion indicates that haloes less massive than  $\sim 5 \times 10^{12}$  are defined by  $\delta(>k_{\text{cut}})$  (in good agreement with the mass-volume defined by  $k_{\text{cut}}$ ) and therefore can be considered independent between realizations. The relative halo mass dispersion can be used as a guide for setting a threshold above which haloes can be studied on a case-by-case ( $M(k_{\text{cut}}) \sim 2 \times 10^{13} h^{-1} M_\odot$  in our case). At lower masses the relative mass dispersion is larger than 30%.

Figure 5 shows a more detailed view of the distribution of halo mass across the ensemble for the 5 most massive haloes in the simulation. The physical location of the five most massive haloes in the simulation box is shown on the left panel and their ensemble mass dispersion is shown in the right panel. The ensemble mass distribution per halo has a roughly log-normal distribution. Note the bump in the mass distribution of the second largest halo. This is the same cluster depicted in Figures 3 and 4. The bump seems to be associated to the large subhalo seen in practically all realizations coming from at the top-right side of the cluster as small variations between realizations make the FoF algorithm link the large subhalo in some cases and not in others.

### 5.3.2 Ensemble Halo Mass Function

Figure 6 shows the mass function computed from 20 individual realizations (gray lines) and the full ensemble (black line). We also show the ratio between the individual realizations and the ensemble in the bottom panel. The individual mass functions closely follows the ensemble mass function with some variations becoming more significant for masses larger than  $10^{13} h^{-1} M_\odot$ , at which point the mass function is dominated by the relatively small volume of the simulation box. The “knee” in the mass function at  $M \sim 10^{13} h^{-1} M_\odot$

is a particular characteristic of the original template used to create the ensemble. The volume-mass defined by  $k_{\text{cut}}$  limits the halo-mass dispersion between realization so we do not expect large variations in mass between large haloes. From the volume-mass defined by  $k_{\text{cut}}$  we expect to have roughly the same halo population above a mass of  $\sim 10^{13} \text{ h}^{-1} \text{ Mpc}$ .

## 6 TWO ILLUSTRATIVE ENSEMBLE STRUCTURES

In this section we present two structures that (given their scale) are persistent across the ensemble and can be stacked into an “ensemble structure”. We focus on two cases of environment extremes: the second most massive cluster in the simulation and a large void next to it. Here we offer a general qualitative description of the halo population in both environments. We believe that every case presented here deserves a more detailed study that goes beyond the scope of this introductory paper. A more detailed analysis will be presented in a forthcoming paper in preparation.

### 6.1 Ensemble cluster

#### 6.1.1 Density field

Figure 8 shows a comparison between a single realization of a cluster (left panel) and the stacked ensemble (right panel). On the top panels we show the density field (saturated at  $\delta > 100$ ) and on the bottom panels the subhalo population on a slice of  $2 \text{ h}^{-1} \text{ Mpc}$  thick. The ensemble field was computed as the mean of the density field computed at each voxel between all realizations.

The density field in the single realization presents several noisy substructures traced by subhaloes and small filaments compared to the more smooth ensemble density field. The averaging across the ensemble has a similar result as “averaging out” the differences in small scales fluctuations between realizations and emphasizing only the common large-scale structures (see also Figure 3). For comparison we show the Lagrangian-smoothed density field at  $\sigma = 2 \text{ h}^{-1} \text{ Mpc}$  in the insert at the top left panel. The Lagrangian-smoothed field has the same large-scale structures than the ensemble average. It even has the large infalling clump at the top-right of the cluster. However, the Lagrangian-smoothed density field lacks of haloes below the mass equivalent to the smoothing scale. The ensemble density field is dominated by features arising from scales above  $k_{\text{cut}}$  but in contrast to the Lagrangian-smoothed field it also contains a large population of collapse objects.

The ensemble density field allows us to remove the randomness in halo properties arising from stochastic processes intrinsic to halo formation (nature) by “averaging out” the small-scale contributions and emphasizing the large-scale environment (nurture). We can now see what is the “true” structure of the cluster. This is the structure defined not by the primordial small-scale fluctuations inside the cluster, but by the large-scale environment of the cluster. The most salient feature of the cluster is the large clump of matter on the top-right side. This clump is formed by large subhaloes present in practically all realizations. The large subhalo is infalling into the main halo through the filament coming from

the top-right. It is important to emphasize that the clump in the stacked ensemble density field does not correspond to an individual halo but to an ensemble of subhaloes. Because of that we can not directly compute the clump’s properties from the density field as in a standard simulation. Instead we would have to identify the subhalo across the ensemble and compute its ensemble moments.

#### 6.1.2 Subhalo population

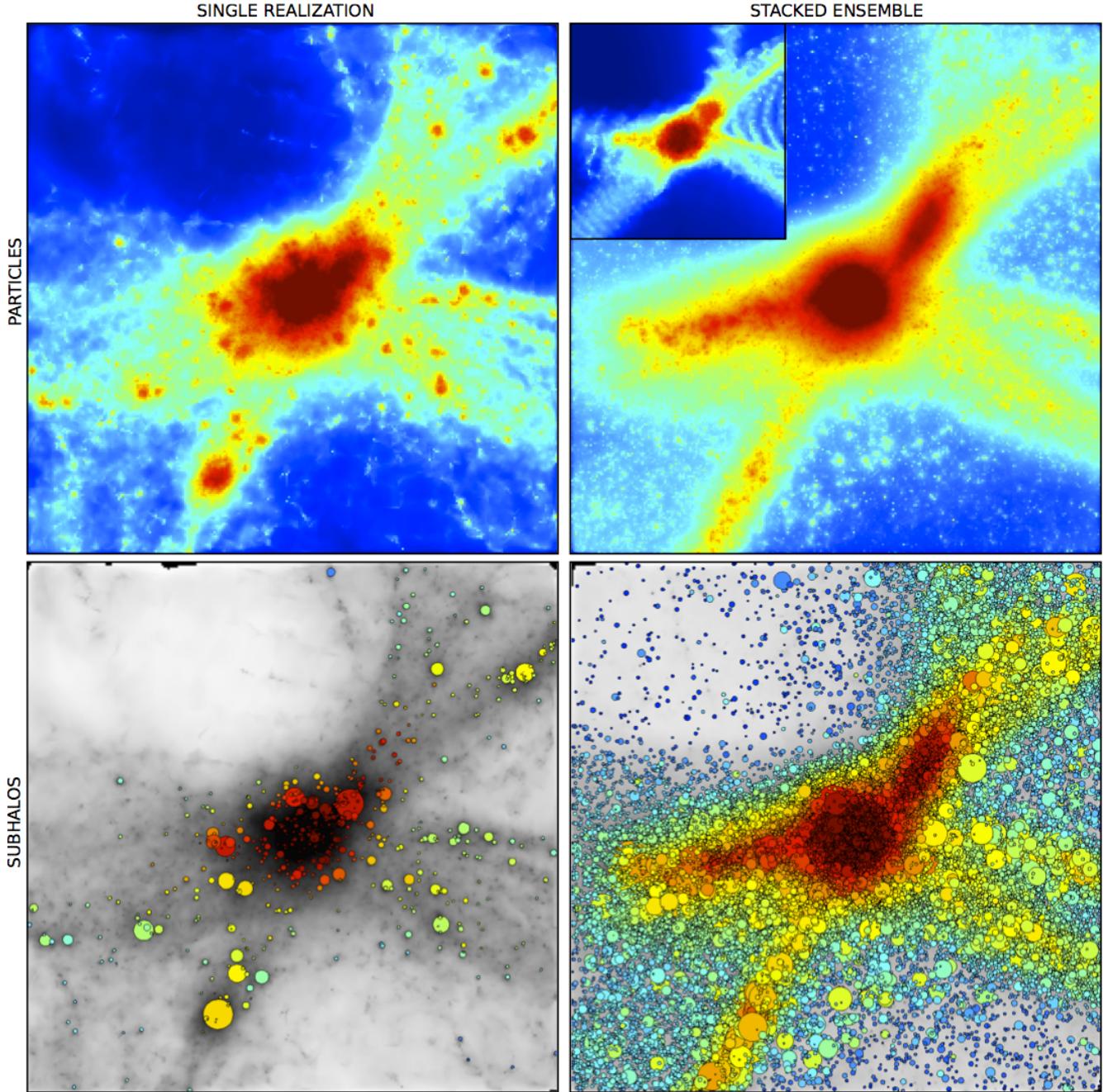
The bottom panels of Figure 8 show the subhalo population inside the second largest halo in the simulation. We included haloes with more than 20 particles corresponding to a mass of  $3.25 \times 10^9 \text{ h}^{-1} \text{ M}_\odot$ . This figure illustrates the advantages of using of the ensemble over single simulations. The subhalo population on the single realization barely samples the central halo and its surrounding structures. Filaments are sampled by haloes but not densely enough to trace the filament’s shape. We can see some dependence between halo size and local density but because of the stochastic nature of halo formation we can not draw any significant conclusions from the single realizations. On the other hand, the ensemble simulation shows an astounding level of detail. We can now clearly see the distribution of haloes/subhaloes inside and around the central halo. Subhaloes in the ensemble fill all the space in and around the cluster. Even the tenuous filaments are densely traced by hundreds of haloes. Consider that for every halo in the single realization we expect in average 220 haloes in the stacked ensemble. Haloes can be seen in all environments, from the dense core of the central cluster down to the center of the underdense adjacent voids. There are  $\sim 5 \times 10^6$  haloes in the ensemble giving a mean halo density of the order of  $\sim 150$  haloes per cubic megaparsec. This value is even higher in high-density regions such as the environment of the cluster in Figure 8 where one can easily achieve number densities of a thousand objects per cubic megaparsec.

### 6.2 Ensemble void

We now focus on a large void adjacent to the central halo in Figure 8. The void can be seen on the top-left side of the central cluster in the same figure. As in the case of the cluster, we describe first the density field and then the halo population.

#### 6.2.1 Density field

The density field from a single realization and the complete ensemble are shown on the top panels of Figure 9. There are significant differences between the single realization and ensemble density fields. The single realization shows more substructure inside the void as a result of its internal small-scale fluctuations  $\delta_{\text{gal}}$ . We can see several tenuous filaments crossing the void in the single realization density field. The stacked ensemble density field on the other hand is very smooth as a result of the averaging over the 220 realizations. The ensemble averaging allows us to highlight the persistent properties of the void given by its surrounding LSS and not its internal random fluctuations. We can now see variations in the density profile of the void. The density profile

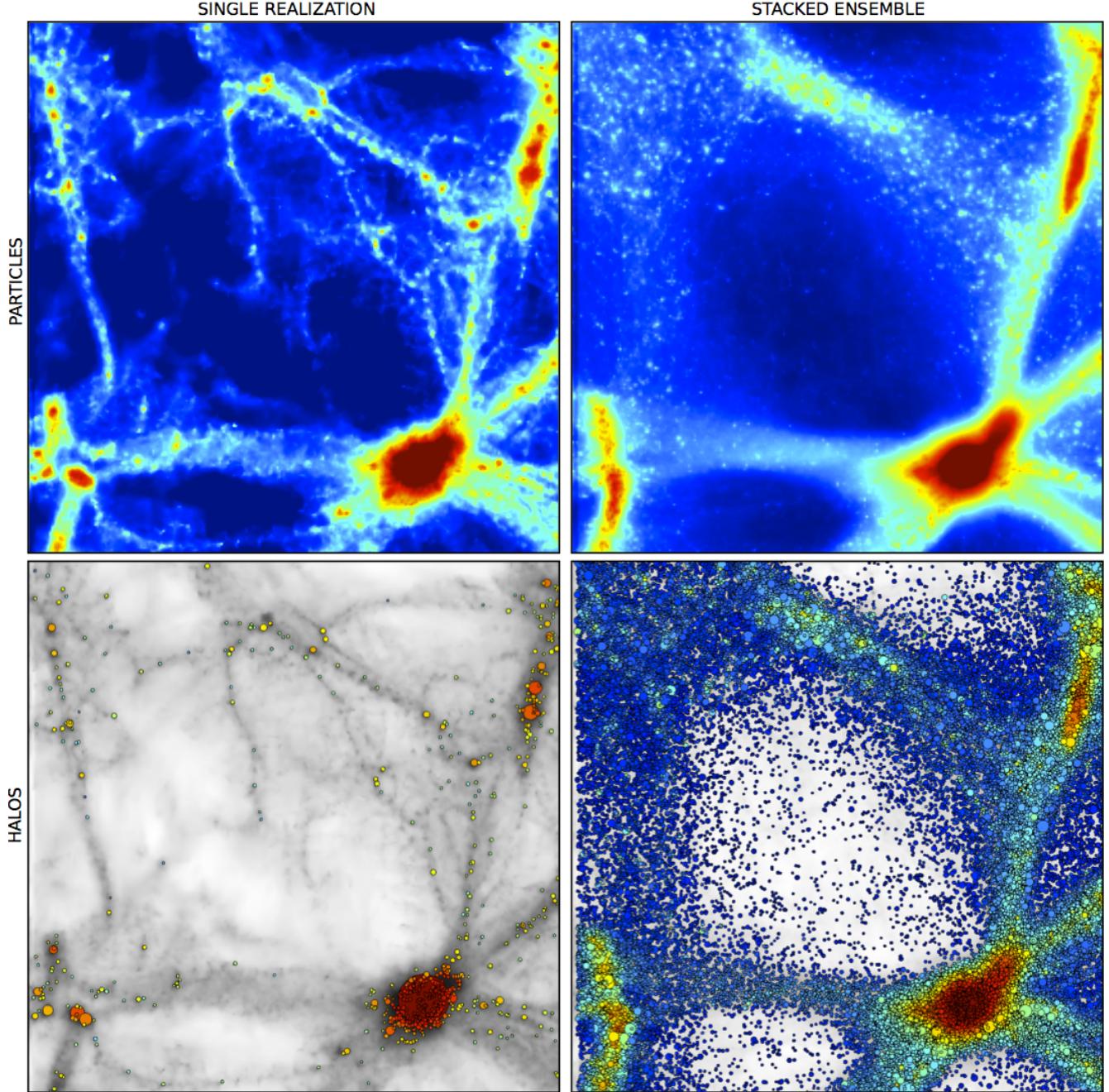


**Figure 8. Single vs. ensemble simulation of a cluster (left and right panels respectively).** Top panels: density across the simulation box on a thin slice of  $8 \text{ h}^{-1}$  Mpc of side and  $2 \text{ h}^{-1}$  Mpc of thickness. The density field from a Lagrangian-smoothed realization with  $\sigma = 2 \text{ h}^{-1}$  Mpc is shown for comparison in the insert of the top-right panel (see text for details). Bottom: subhaloes identified inside the slice. The circles are scaled with the halo radius and colored with the value of the ensemble density field from the top right panel. For clarity we only show subhaloes with radius smaller than  $500 \text{ h}^{-1}$  Kpc. The gray background corresponds to the density field for one single realization and the ensemble (left ad right panels respectively).

in the top-left to bottom-right direction is steeper than in the bottom-left to top-right direction. This difference is already hinted in the single realization density field but it is not possible to determine if it is an intrinsic property of the void or the result of its internal random fluctuations.

#### 6.2.2 Halo distribution

The distribution of FoF haloes is shown in the bottom panels of Figure 9. Here we plot the FoF haloes instead of SUBFIND subhaloes as done in the case of the cluster (see Figure 8) in order to illustrate how haloes trace the internal structure of the void and not the internal structure of haloes themselves. The difference between single and ensemble simulations in voids is even more striking than in the case of



**Figure 9. Single vs. ensemble simulation of a void (left and right panels respectively).** Top panels: density across the simulation box on a thin slice of  $16 \text{ h}^{-1} \text{ Mpc}$  of side and  $2 \text{ h}^{-1} \text{ Mpc}$  of thickness. Bottom: FoF haloes identified inside the slice. The circles are scaled with the halo radius and colored with the value of the ensemble density field from the top right panel. The gray background corresponds to the density field for one single realization and the ensemble (left ad right panels respectively).

clusters (see Figure 8). Voids are practically devoid of galaxies at the minimum halo mass of our simulation ( $3.25 \times 10^9 \text{ h}^{-1} \text{ M}_\odot$ ). We can see a tenuous filament going from the top of the cluster down to its center where it seems to disappear as the filament continues beyond the slice. From the single realization it is not possible to determine if the filament is the result of the large-scale environment or smaller local fluctuations. The extremely low halo density also prevents us from computing local halo densities which would be zero for a “fair” density estimation or completely dominated by

discreteness effects. Given the low or even null halo density it is not possible to compute any meaningful halo statistics inside the void. This is the typical case of a void with its intrinsic “observational” limitations. The finite halo density prevents us from doing any type of local analysis. On the other hand, the ensemble simulation gives a completely different view of the same void. Haloes cover the entire volume of the void even down at its extremely underdense center. The stacking of realizations increases the halo density by a factor of 220 and the signal-to-noise increases to  $s/n = 15$ .

By increasing the number density of haloes are able to compute local densities even in the most underdense environments where one would normally be dominated by noise. As we will see in the following section we are able to compute even more elaborated local statistics such as local mass functions even at the center of voids.

## 7 MASS DEPENDENCE ON HALO BIASING

Figure 10 shows the tracing of the Cosmic Web by haloes in two mass ranges:  $10^{10} - 10^{11} \text{ h}^{-1}$  and  $10^{11} - 10^{12} \text{ h}^{-1} \text{ M}_\odot$  (left and right panels) for the single and stacked ensembles (bottom and middle panels). This illustrates in a unique way the biasing of haloes with respect to the underlying dark matter density field (Kaiser 1984). Halo biasing depends on many factors such as mass, scale and halo properties (Mo & White (1996); Wechsler et al. (2006); Smith et al. (2007) among others). In general, the limited halo density in standard realizations makes it impossible to study the biasing at particular environment. The unprecedented specific halo density of the MIP simulation allows us to compute matter and halo densities in all environments and over a wide halo mass range.

The difference in the halo distribution between the single and stacked ensembles is striking for both halo mass ranges. Haloes in the stacked ensemble closely trace the LSS in contrast with the single-realization where the traced LSS looks shallow and dominated by sampling noise. Halo biasing has the effect of sharpening the features in the LSS, this effect that can be exploited to select LSS structures in complex environments (Aragon-Calvo in preparation). This sharpening effect can not be solely explained by local density as we will see in the following section.

### 7.1 Bias-induced LSS sharpening

Figure 11 shows a dramatic example of the bias-induced sharpening of LSS structures. The figure highlights a filament extending between two clusters. This is the same kind of straight filaments found by Colberg et al. (2005) and Dolag et al. (2006). The high halo density combined with their biased distribution sharpens the boundaries of the internal structure of the filament unveiling its complex structure. This filament is in fact formed by three different filaments collapsing along their mutual axis. The face-on view of the filament in the right panel of Figure 11 clearly shows three individual filaments with similar thickness. This kind of structures are only observed in stacked ensembles where the small-scale fluctuations are averaged-out and the high halo density allows for the local effects of halo biasing to be observed.

The distribution of high-mass haloes (top-right panel) is an excellent illustration of the effect of environment on local biasing. The large void at the center of the slice does not contain any halo even after 220 realizations. Note the sharp boundaries of the void, a small change in the local density translates into a sharp fall in the halo density profile. Compare the matter (gray line) and stacked ensemble halo (black solid line) density profiles shown in the bottom panels. This local biasing effect can not be observed in the

single-realization case where there are not enough haloes in this mass range to trace the LSS.

The density profiles shown in the bottom panel demonstrate the superior halo sampling of the stacked ensemble compared to the single realization. The density field was reconstructed from the halo distribution using the DTFE technique (Schaap 2007). Voids traced by haloes in the single realization are shallower than in the stacked ensemble where the cosmic structures have sharp boundaries. The low density of haloes in the single-realization makes it impossible to draw any conclusions on the behaviour of the local biasing.

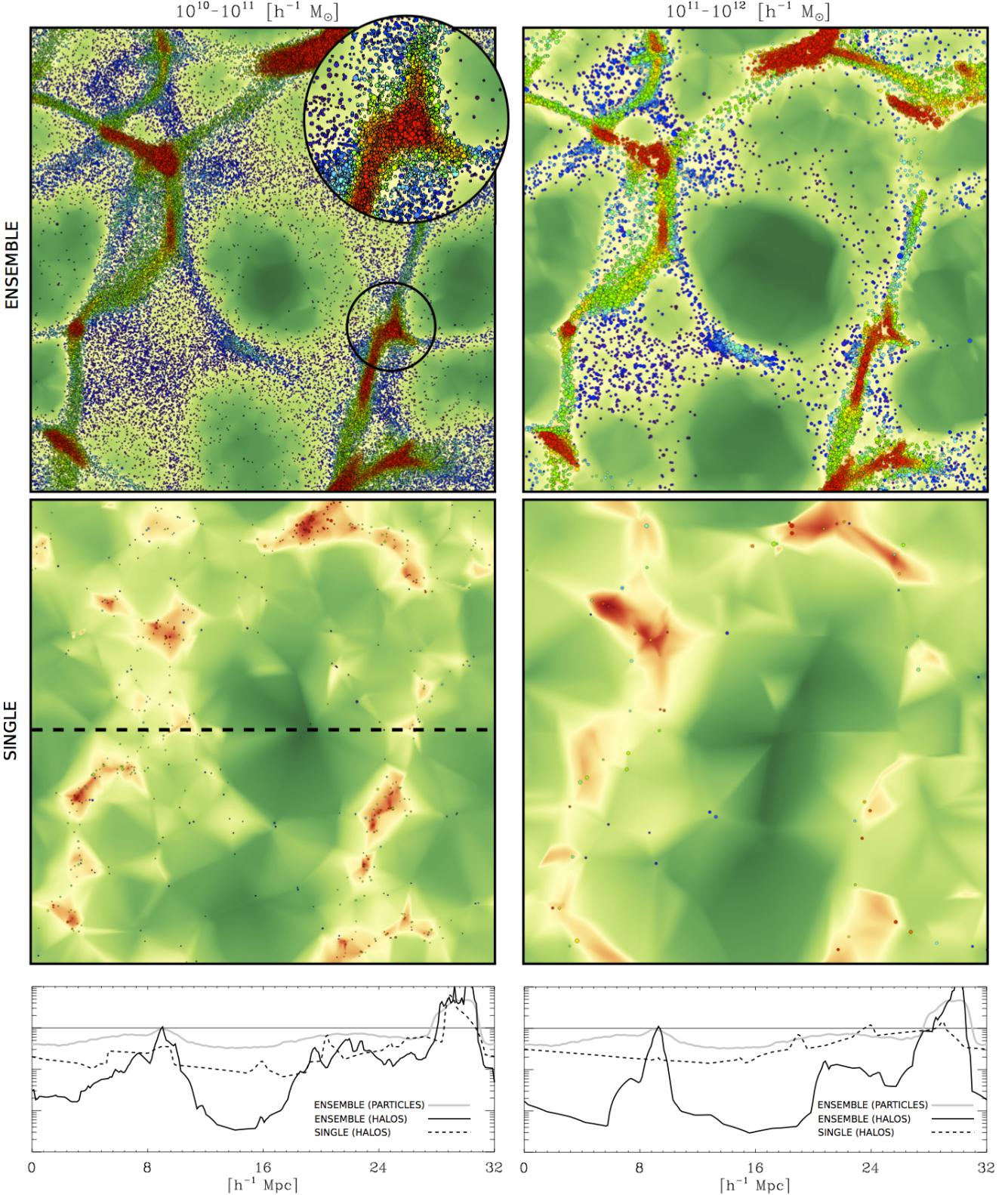
## 8 LOCAL HALO MASS FUNCTION

The unprecedented local halo density of the MIP simulation allows us to perform local analysis that are in general computed taking advantage of ergodicity, exchanging volume averages over ensemble averages. In this section we present a direct application of local ensemble statistics by computing the *local halo mass function*. This is fundamentally different from previous studies based on global or environment-based halo samples. The analysis presented here is performed on a point basis.

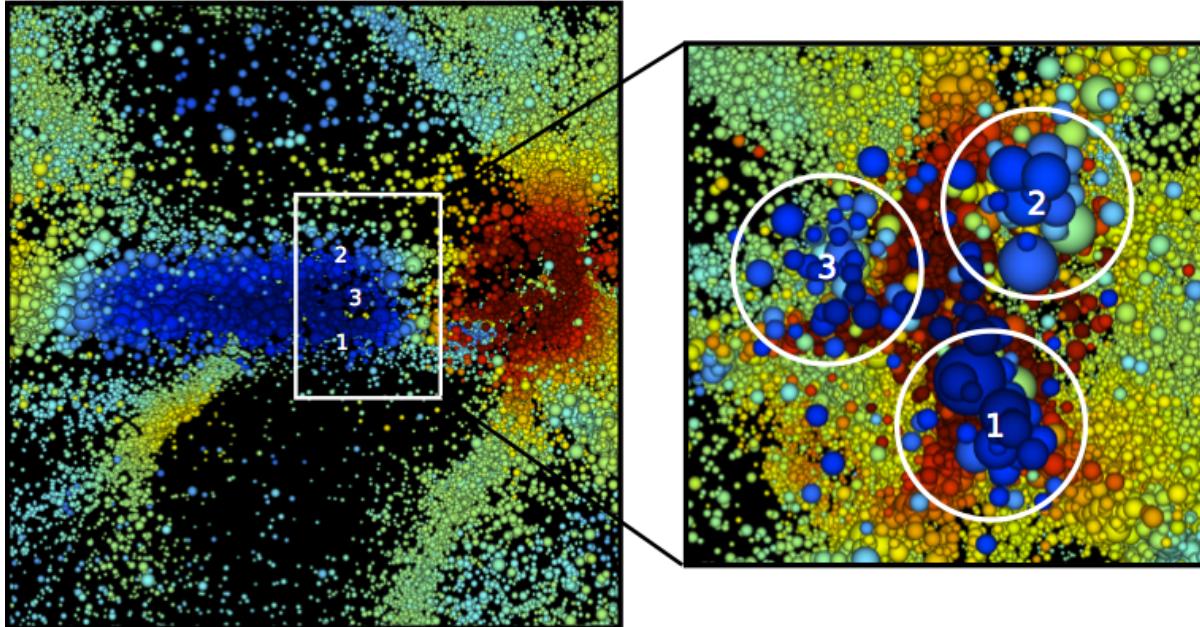
Figure 12 shows a slice of the density field (left panel) and four spherical windows sampling the four basic LSS environments: void, wall, filament and cluster. We also sample a linear path at 13 positions starting at the center of a void and ending at the center of a filament in the boundary of the void. The halo mass function was computed from all the haloes in the ensemble located inside each sampling window. The mass function inside each sampling window is shown in the bottom-left panels of Figure 12. The thick gray line corresponds to the ensemble global mass function (see also Figure 6). Sampling inside a void, wall, filament and cluster are indicated in both panels by solid, dashed, dotted and dashed-dotted lines respectively.

There is a clear trend in the height, slope and characteristic mass of the mass function with LSS environment. Both voids and walls have mass functions lower than the global mass function. The mass function in filaments and clusters is higher than the global mass function. The slope of the mass function is also higher for voids compared to other environments (see Gottlöber et al. (2003)). We find that the mass function inside the selected filament is basically an enhanced version of the global mass function. The mass function at the center of the cluster is also higher and its slope is also steeper than the global mass function. Around a mass of  $\sim 10^{12} \text{ h}^{-1} \text{ M}_\odot$  the mass function in the cluster and the filament cross and after that point there are more massive haloes in the filament than in the cluster. This lack of medium mass haloes may be the result of the haloes being inside the cluster and therefore not accounted for in the FoF catalogue. A more detailed study will be presented in a forthcoming paper. The mass function inside the wall is a fraction of the general mass function but it rapidly drops after  $\sim 3 \times 10^{11} \text{ h}^{-1} \text{ M}_\odot$ . For this particular wall the probability of hosting a Milky-Way halo is lower than  $10^{-6}$ . Even after 220 realizations there are no haloes in that mass range at this particular position.

The lower-right panel of Figure 12 shows the local ensemble mass function sampled along a linear path starting at the center of a void and ending at a small filament. The



**Figure 10.** Comparison of the spatial distribution of haloes as function of their mass between a single realization (middle panels) and the stacked ensemble (top panels). Two mass ranges are shown in the left and right panels. Haloes are indicated by circles scaled with their radius and colored according to their local ensemble density. A detailed view of the halo distribution is shown in the background with green-red color table. Note that there are two color tables shown in the figure. One for the haloes and another for the reconstructed density field. The bottom panels show the density profile sampled across the center of the slice indicated by the dashed line on bottom-left panel. Gray lines correspond to the ensemble dark matter density profile. Solid and dashed lines correspond to single and ensemble halo density profiles respectively.



**Figure 11.** A filament composed by three merging filaments. The left panel shows a 3D view of a large cluster (red) and a filament next to it (blue). The spheres correspond to subhaloes in the stacked ensemble scaled with their radius and colored according to their local dynamics (Aragon-Calvo in preparation). The three filaments are indicated by numbers. The right panel shows a “face-on” view of the filaments in the volume highlighted in the left panel. We draw circles to indicate the location of the three filaments forming a “cosmic fork”.

local density increases away from the center of the void. The value of the smoothed ( $\sigma_G = 2 \text{ h}^{-1} \text{ Mpc}$ ) density fluctuations at the two end ends of the path are  $\delta_{2 \text{ Mpc}} = -0.51$  and  $\delta_{2 \text{ Mpc}} = 0.31$  respectively. The corresponding raw density fluctuation values are  $\delta = -0.82$  and  $\delta = 2.12$ . This is reflected in the local mass function which also increases its height and decreases its slope as we move to denser regions. Note how the mass function changes uniformly across the sampling points in the void in contrast to the significant differences between environments. The mass function smoothly increases indicating a smooth change in the environment.

In order to quantitatively compare the local mass function at several sampling positions we parametrized the halo mass function using a Schechter function:

$$\phi = \frac{\phi_*}{M_*} \left( \frac{M}{M_*} \right)^\alpha \exp \left( -\frac{M}{M_*} \right), \quad (14)$$

where  $\phi_*$  determines the height of the curve,  $\alpha$  is the slope of the power-law scaling in the first part of the curve and  $M_*$  is the characteristic mass defining the transition point to an exponential behavior.

The mapping between the linear path shown in the top-left panel of Figure 12 and the  $M_* - \alpha$  parameter space is shown in the top-right panel of Figure 12. As the sampling window moves from the center of the void up to the dense filament the characteristic mass increases and the absolute magnitude of the slope  $\alpha$  decreases. Note that the linear path in configuration space maps into a relatively smooth path in the  $M_* - \alpha$  parameter space. This indicates that in low density regions not only one expect less haloes but also the ratio between more and less massive haloes decreases. A simple scaling of the mass function with local density can not explain the change in the slope and characteristic

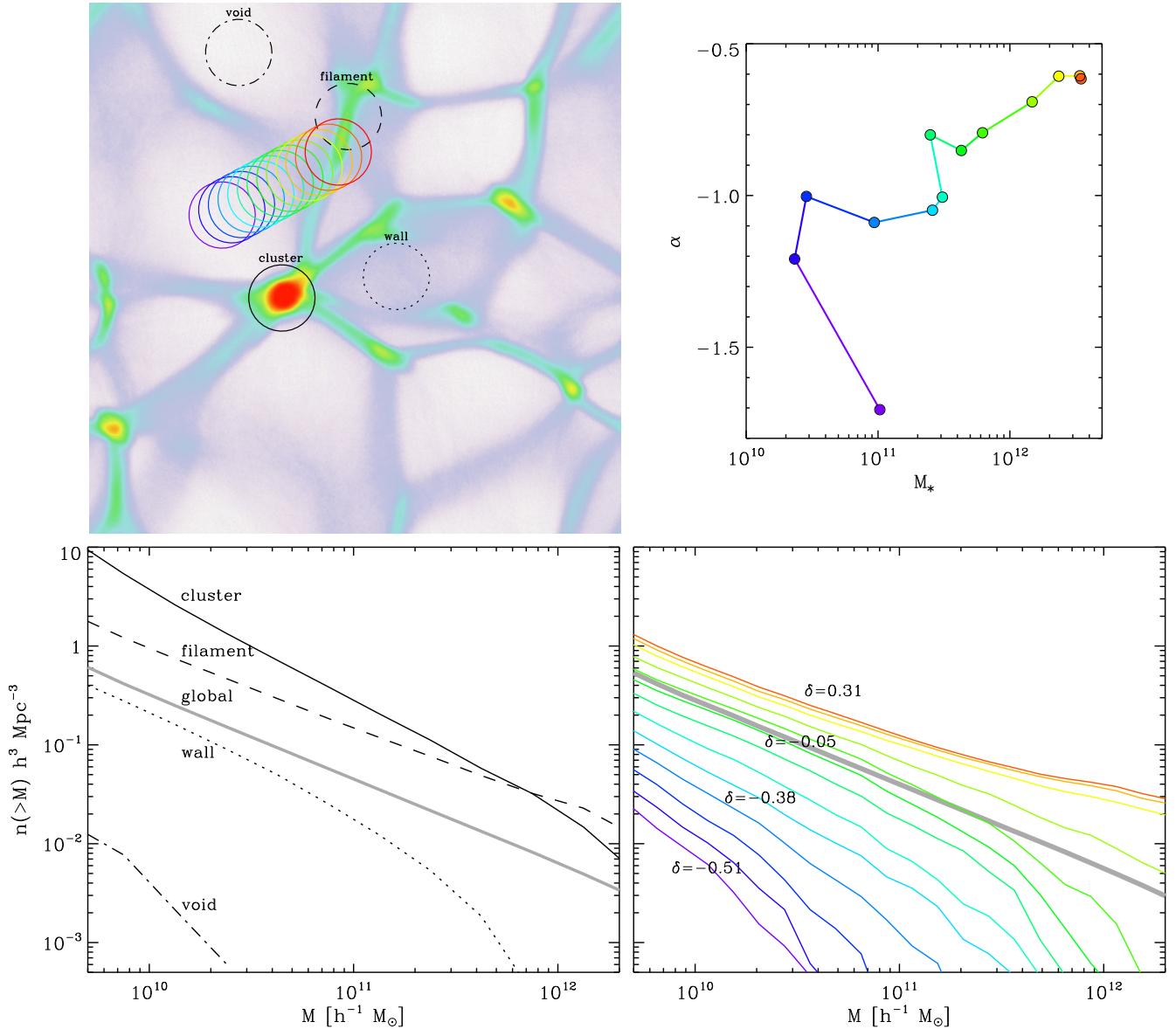
mass of the mass function. The increase in the slope of the mass function is a strong indication that halo formation is suppressed inside voids due to other processes in addition to local density.

We would like to emphasize that the results presented here correspond to the local ensemble statistics at the particular sampling positions, not just global trends. We can see continuous changes in the mass function as we move through the simulation box. This is a significant advance with respect to previous studies. Here we can, for the first time, establish a direct link between the halo population and the particular properties of the LSS environment at any given position.

## 9 CONCLUSIONS AND FUTURE PROSPECTS

We introduced a novel application of constrained ensembles where all realizations share the same large-scale fluctuations while having independent small-scale fluctuations. This technique represents a significant improvement over standard single-realization N-body simulations. By creating a large set of semi-independent realizations and stacking them we break the fundamental limit of specific halo density and are able to trace all LSS environments in great detail, from the dense clusters down to the underdense center of voids.

We presented the *Multim In Parvo* (MIP) simulation/project consisting on a constrained ensemble simulation with 512 realizations from which 220 have been fully ran and analyzed. The MIP simulation in its current status is equivalent in terms of volume and number of particles to a  $193 \text{ h}^{-1} \text{ Mpc}$  box containing  $\sim 1540^3$  particles and  $\sim 5 \times 10^6$  haloes with a minimum halo mass of  $3.25 \times 10^9$



**Figure 12. Local ensemble halo mass function** The left panel shows a slice of the density field and several sampling windows corresponding to the four basic LSS environments: cluster, filament, wall and void (solid, dashed, dotted and dotted dashed lines respectively) and 13 samplings going from the center of a void to a filament at its boundary (purple to red color lines). The bottom-left panel shows the corresponding local halo mass function computed from the halo ensemble at the four environments. The bottom-right panel shows the local mass function computed along the 13 sampling points shown in the top-left panel. We indicate the density field smoothed at  $\sigma = 2 \text{ h}^{-1} \text{ Mpc}$  at the center of each sampling window. The top-right panel shows the effect of the local environment on the mass function parametrized by its slope  $\alpha$  and characteristic mass  $M_*$  (see text for details). The colored filled circles joined by lines correspond to the sampling path shown in the top-left panel.

$\text{h}^{-1} M_\odot$ . As a comparison, the unprecedented *halo* density in the MIP simulation is larger than the *particle* density of the millennium simulation by almost a factor of 2!

We showed preliminary results on the mass dependence on the halo biasing and how this effect can be used to identify structures in complex environments.

The unprecedented specific halo density of the MIP simulation allows us to compute ensemble statistics on a local basis. We presented the first examples of the *local mass function* computed at several representative LSS environments. We also computed the local halo mass function along a path through the simulation box and showed the mapping of the

path in configuration space to a path in the  $M_* - \alpha$  parameter space.

### 9.1 Future prospects

Given its broad scope this technique can be applied to a variety of physical phenomena over a wide range of scales. One promising application is the use of stacked constrained ensembles of the local supercluster to compute local ensemble statistics of our own cosmic neighborhood and use this information to interpret local observations.

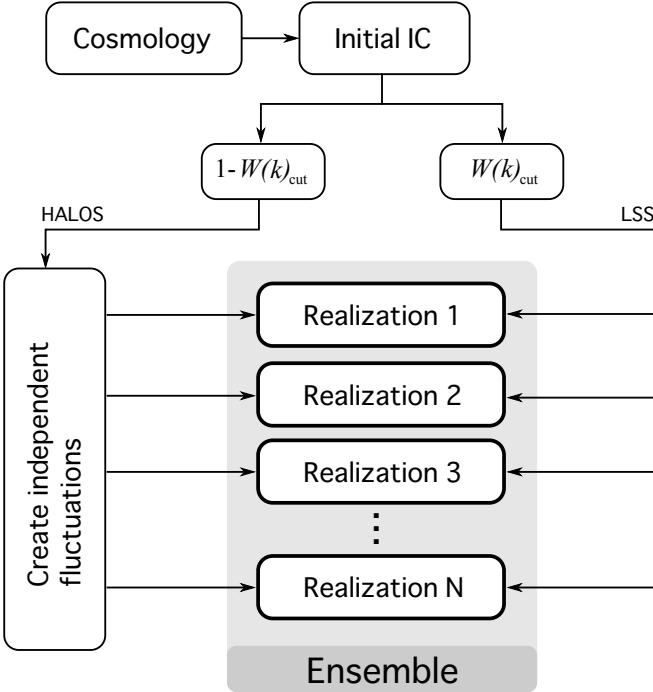
In a following series of papers we will explore the effect of the LSS on halo properties and the ensemble properties of large groups and clusters as function of their LSS environment.

## 10 ACKNOWLEDGEMENTS

This research was funded by the Gordon and Betty Moore foundation. I would like to thank Mark Neyrinck, Bernard Jones and Joe Silk for many stimulating discussions and proofreading of the early manuscript.

## REFERENCES

- Angulo, R. E., Springel, V., White, S. D. M., et al. 2012, arXiv:1203.3216
- Aragón-Calvo, M. A., van de Weygaert, R., Jones, B. J. T., & van der Hulst, J. M. 2007, ApJL, 655, L5
- Aragón-Calvo, M. A., Jones, B. J. T., van de Weygaert, R., & van der Hulst, J. M. 2007, A&A, 474, 315
- Aragón-Calvo, M. A., Platen, E., van de Weygaert, R., & Szalay, A. S. 2010, ApJ, 723, 364
- Aragon-Calvo, M. A., van de Weygaert, R., Araya-Melo, P. A., Platen, E., & Szalay, A. S. 2010, MNRAS, 404, L89
- Aragon-Calvo, M. A., Silk, J., & Szalay, A. S. 2011, MNRAS, 415, L16
- Aragon-Calvo, M. A. & Szalay, A. S. 2012, MNRAS, in press.
- Avila-Reese, V., Colín, P., Valenzuela, O., D’Onghia, E., & Firmani, C. 2001, ApJ, 559, 516
- Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15
- Byrd, G., & Valtonen, M. 1990, ApJ, 350, 89
- van den Bosch, F. C. 2002, MNRAS, 331, 98
- Cautun, M., van de Weygaert, R., & Jones, B. J. T. 2012, arXiv:1209.2043
- Colberg, J. M., Sheth, R. K., Diaferio, A., Gao, L., & Yoshida, N. 2005, MNRAS, 360, 216
- Lacey, C., & Cole, S. 1993, MNRAS, 262, 627
- Diemand, J., Kuhlen, M., & Madau, P. 2007, ApJ, 657, 262
- Dolag K., Meneghetti M., Moscardini L., Rasia E., Bonaldi A., 2006, MNRAS, 370, 656
- Einasto, J., Hütsi, G., Saar, E., et al. 2011, A&A, 531, A75
- Forero-Romero, J. E., Hoffman, Y., Yepes, G., et al. 2011, MNRAS, 417, 1434
- Gao, L., Navarro, J. F., Frenk, C. S., et al. 2012, MNRAS, 425, 2169
- Gottlöber, S., Klypin, A., & Kravtsov, A. V. 2001, ApJ, 546, 223
- Gottlöber, S., Lokas, E. L., Klypin, A., & Hoffman, Y. 2003, MNRAS, 344, 715
- Hahn, O., Porciani, C., Carollo, C. M., & Dekel, A. 2007, MNRAS, 375, 489
- Kaiser, N. 1984, ApJL, 284, L9
- Klypin, A., Kravtsov, A. V., Valenzuela, O., & Prada, F. 1999, ApJ, 522, 82
- Libeskind, N. I., Hoffman, Y., Knebe, A., et al. 2012, MNRAS, 421, L137
- Little, B., Weinberg, D. H., & Park, C. 1991, MNRAS, 253, 295
- Mecke, K. R., Buchert, T., & Wagner, H. 1994, A&A, 288, 697
- Melott, A. L., & Shandarin, S. F. 1993, ApJ, 410, 469
- Meiksin, A., & White, M. 1999, MNRAS, 308, 1179
- Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347
- Orban, C. 2012, arXiv:1201.2082
- Press, W. H., & Schechter, P. 1974, ApJ, 187, 425
- Peebles, P. J. E. 1980, Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435
- Peebles, P. J. E. & Nusser, A., 2010, Nature, 465, 565–569
- Sandage, A., Tammann, G. A., & Hardy, E. 1972, ApJ, 172, 253
- Sandage, A. 1986, ApJ, 307, 1
- Schaap, W. E. The Delaunay Tessellation Field Estimator, 2007, PhD Thesis, University of Groningen
- Schmalzing, J., Buchert, T., Melott, A. L., et al. 1999, ApJ, 526, 568
- Schneider, M. D., Cole, S., Frenk, C. S., & Szapudi, I. 2011, ApJ, 737, 11
- Smith, R. E., Scoccimarro, R., & Sheth, R. K. 2007, Phys. Rev. D, 75, 063512
- Sheth, R. K. 2005, MNRAS, 364, 796
- Smith, R. E. 2009, MNRAS, 400, 851
- Sousbie, T., Pichon, C., Colombi, S., Novikov, D., & Pogosyan, D. 2008, MNRAS, 383, 1655
- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629
- Springel, V., Wang, J., Vogelsberger, M., et al. 2008, MNRAS, 391, 1685
- Suhhonenko, I., Einasto, J., Liivamägi, L. J., et al. 2011, AAP, 531, A149
- Szapudi, I. 1998, ApJ, 497, 16
- Takahashi, R., Yoshida, N., Takada, M., et al. 2009, ApJ, 700, 479
- Tully, R. B., Shaya, E. J., Karachentsev, I. D., et al. 2008, ApJ, 676, 184
- Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, ApJ, 652, 71
- White, S. D. M. 1984, ApJ, 286, 38
- Hao-Yi Wu, Hahn, O., Wechsler, R. H., Mao, Yao-Yuan, Behroozi, P. S. 2012, astroph-1209.3309
- Zabludoff, A. I., & Mulchaey, J. S. 1998, ApJL, 498, L5
- Zhang, Y., Yang, X., Faltenbacher, A., et al. 2009, ApJ, 706, 747
- Zel'dovich Ya. B., 1970, A&A, 5, 84
- Doroshkevich, A. G., Kotok, E. V., Poliudov, A. N., et al. 1980, MNRAS, 192, 321
- Einasto, J., Joeveer, M., & Saar, E. 1980, MNRAS, 193, 353
- Einasto, J., Einasto, M., & Gramann, M. 1989, MNRAS, 238, 155
- Einasto, J., Einasto, M., Gottlöber, S., et al. 1997, Nature, 385, 139
- Einasto, J., Suhhonenko, I., Hütsi, G., et al. 2011, A&A, 534, A128
- Joeveer, M., & Einasto, J. 1978, Large Scale Structures in the Universe, 79, 241
- Stanonik, K., Platen, E., Aragón-Calvo, M. A., et al. 2009, ApJL, 696, L6



**Figure A1.** Block diagram of the construction of the ensemble simulation.

#### APPENDIX A: ENSEMBLE GENERATION

Figure A1 shows the steps involved in the creation of the ensemble simulation. We start by defining the cosmological model, box size and particle resolution. From that information We then create the template initial conditions file (IC). The field is then separated into large and small-scale fluctuations using the sharp- $k$  filter defined in equation 4. The small-scale fluctuations are shifted along each dimension in configuration space by a physical scale corresponding to  $k_{\text{cut}} = 4 \text{ h}^{-1} \text{ Mpc}$ , producing a total of  $8 \times 8 \times 8 = 512$  independent realizations (another possibility is to generate completely independent random small scale fluctuations per realization). The large-scale fluctuations are simply replicated for each realization in the ensemble. Each initial conditions file is the evolved using the Zel'dovich formalism and run.